

Chapter 33: On-Farm Research and Analysis



C.G. Carlson (Gregg.Carlson@sdstate.edu)
Cuirong Ren
D.E. Clay (David.Clay@sdstate.edu)

This chapter provides a primer on analyzing findings from on-farm research studies. This chapter complements Chapter 32. In many on-farm studies, treatments are implemented across landscapes (Fig. 33.1). To make informed, science-based decisions, understanding data collection and simple arithmetic calculations used for statistical analysis and data interpretation are critical. Microsoft Excel is used in the following discussion.

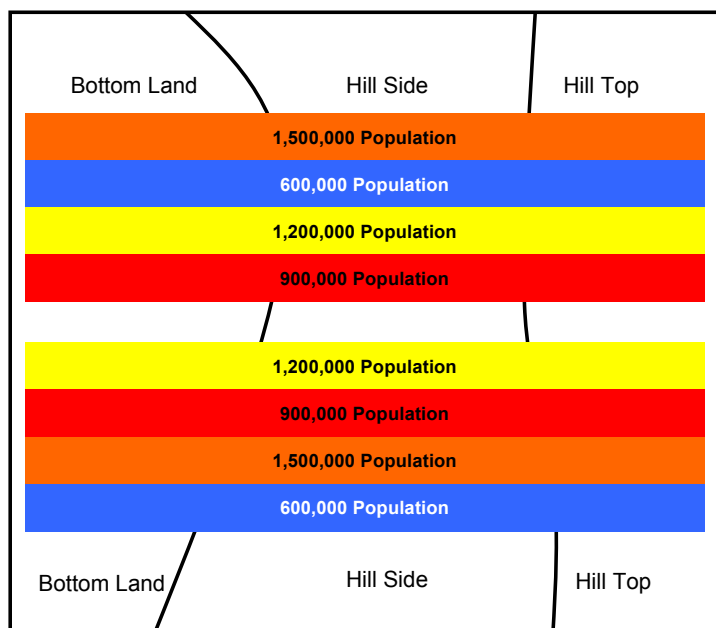


Figure 33.1. On-farm research study, where treatments are placed perpendicular to elevation. (Carlson et al. 2011)

Loading the software into Microsoft Excel

For this section, you will need to load software into your Microsoft Excel. To load these programs, follow the discussion below. Different protocols are needed for loading the required software in Microsoft Excel 1997-2003 and 2007.

In Microsoft Excel 97-03:

1. Use Excel's Tools menu, Data Analysis (Fig. 33.2).

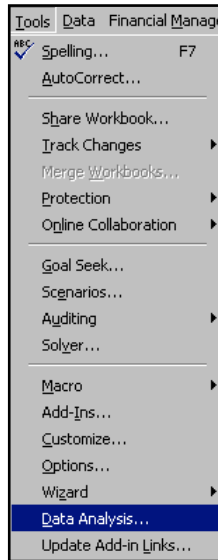


Figure 33.2. Step 1 in loading required software in Excel 97-2003.

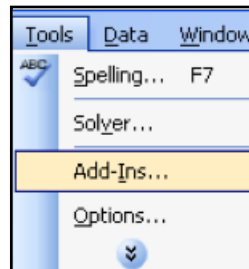


Figure 33.3. Step 2 in loading required software.

2. If the above Data Analysis menu is not available, you will need to turn it on (Fig. 33.3). To do this, select Tools and highlight Add-Ins as below.
3. From the Add-Ins menu, select with a check Analysis ToolPak as shown in Figure 33.4. Now the Data Analysis line will show up from the Tools menu.

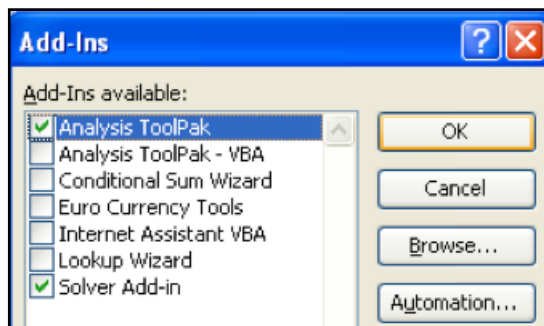


Figure 33.4. Step 3 in loading required software.

4. From the Data Analysis menu, Anova: Two-Factor Without Replication is selected and OK is pushed.

In Microsoft 2007, use the following protocols:

1. Select the Office (File) Button, left hand corner.
2. Select Excel options.
3. Select Add-Ins.
4. Select analysis ToolPak.
5. Select Go.
6. Click Analysis ToolPak.
7. Click OK.

Introduction

On-farm studies can be used to test product claims. For example, at a local farm show, Mel met a dynamic salesman, Lioekans, who was selling a new superior yield enhancing product (this miracle product is sold under the trade name “Super Soak”).

Lioekans sells this incredible product for \$80.00/gallon. Since Mel showed some interest, Lioekans makes a farm visit to Mel’s place. Lioekans suggested that Mel try a free gallon of Super Soak on a single ten-acre strip (suggested application rate is 12.8 ounce/acre) in his field. If Mel gets a yield advantage as advertised, he will agree to use Super Soak on his entire farm next year. If there is no advantage to Super Soak, then the \$80.00 valued product that Lioekans gave Mel is absolutely free without obligation.

Before we go on, let us ask an important question. If Super Soak has zero impact on yield, what is the probability of one untreated strip “A” out yielding a second untreated strip “B”? The answer is 50%. Now, what is the random chance that Super Soak out-yielded the non-treated area if two sets of treated and non-treated strips are used? The answer is 25% ($0.5 \cdot 0.5$). How about if the material is applied to three sets of strips? The answer is 12.5% ($0.5 \cdot 0.5 \cdot 0.5$).

After listening to Lioekans’s sales pitch, Mel agreed to try the product, but with the stipulation that Mel would test the product using four randomly placed replicated strips. Additionally, before Mel agreed to buy the product, the treated plots had to show a statistically significant higher yield at the 95% probability level. With these criteria in place, Lioekans agreed to the terms, drove out of the yard, and was never seen again.

To improve agricultural management, new techniques are frequently created and compared with traditional techniques. On-farm experiments are one approach for comparing the practices on your farm.

On-farm research

Through on-farm research, the influence of a treatment(s) on a measurable variable is investigated. In these experiments, the independent variables are manipulated, whereas, dependent variables are measured. For example, in an N rate experiment where different N rates are applied, the independent variable is N rate and dependent variable is yield.

In an erosion study where the influence of cultivation on erosion is measured, the independent variable might be the number of times that a field is cultivated, while the dependent variable is the amount of erosion. Research experiments are designed to provide insights into cause and effect.

Most experiments contain replications of each treatment. Replications provide assurance that the response to a treatment is real and repeatable and not due to error, variance, or random chance. For example, if you flip a coin once and get heads, with no replication, you may predict that heads will always occur. As more replications of the flip occur, you should see that heads comes up 50% of the time and tails comes up the other 50% of the time.

To implement on-farm studies, field trials should be placed perpendicular to the landscape (Fig. 33.1). To detect differences, it is desirable to maximize the number of replicates. The minimum number of replicates

is three. Additional details about this approach are available at [http://www.ipni.net/ppiweb/ppibase.nsf/\\$webindex/article=3DD9EB2C85256966006141B759745B04](http://www.ipni.net/ppiweb/ppibase.nsf/$webindex/article=3DD9EB2C85256966006141B759745B04)

Statistics overview

Statistics are used to objectively evaluate numerical data in order to make informed decisions. Statistics involve computation and arithmetic manipulation of data that can be *qualitative* or *quantitative*.

Qualitative variables have values that cannot be defined numerically (e.g., sex, plant species, or marital status), whereas, quantitative variables can be defined numerically (e.g., yield, weed density, weight).

Means or averages

The mean, or average, of a sequence of numbers is the sum divided by the number of measurements taken. A notation for the mean or average is \bar{X} . Mathematically, this is expressed by the equation,

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{15}{4} = 3.75$$

This equation indicates that all values of X from the first to the last (n^{th}) will be summed and divided by the total number of observations (n). In Microsoft Excel, the average of numbers can be determined using the command, **=average(start list: end list)**.

Variance and standard deviations

Unless each observation is the same number, there is variation around the average value. The variance (s^2) provides a measure of this variation. The *variance* is a measure of precision (how close the numbers are to each other), not accuracy (how close the numbers are to the true value). The variance (s^2) is calculated using the equation,

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

where, \bar{X} is the mean or average of the measurements, X_i are the individual measurements, and n is the number of measurements taken.

$$\begin{aligned}\bar{X} &= \frac{1+5+6+3}{4} = \frac{15}{4} = 3.75 \\ s^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} = \frac{(1-3.75)^2 + (5-3.75)^2 + (6-3.75)^2 + (3-3.75)^2}{4 - 1} = 4.92\end{aligned}$$

In this calculation, Σ tells you to sum the values where X_i goes from 1 to n . Using the data set where $X_1=1$, $X_2=5$, $X_3=6$, and $X_4=3$, the variance is calculated as follows:

The standard deviation (s) is defined as the square root of the variance (s^2). In this example the standard deviation is 2.22 ($\sqrt{4.92}$). In Microsoft Excel, these terms can be calculated using the commands, **=var(start of list: end of list)** and **=stdev(start of list: end of list)**.

T-test

The t-test provides a statistical comparison between means (\bar{X}) of two normal populations with unknown, but assumed equal, variances. Further reading on this subject is suggested at http://www.statstutorials.com/EXCEL/EXCEL_TTEST1.html. The ‘tails’ of the t-test refer to the extremities of the normal distribution curve, that is, one-tailed or two-tailed tests. Calculated t values are compared to t values found in a Table 33.1 to determine significance of results. Subscripts 1 and 2 refer to the two populations being comparatively tested.

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

Example 1. Determining the composite soil sampling requirements

Subsamples are the number of unique counts or measurements made within a management zone. For example, 15 soil cores are composited into each sample from a management zone. Based on these 15 subsamples, a single fertilizer recommendation is determined for this zone. The accuracy of this recommendation could be improved by increasing the number of samples contained in the composite sample. The number of samples needed for a particular field (sampling requirement) is determined using the following equation.

$$n = \frac{t^2 \cdot s^2}{D^2}$$

In this example, n is the sub-sampling requirement, s is the standard deviation described above, t is the t value, and D is the desired confidence interval. The t value in the sub-sampling requirement equation comes from the t-table shown in Table 33.1.

A more complete discussion of t-tables is available at <http://www.statsoft.com/textbook/sttable.html>. Most t-tables have an “ α ” (pronounced “alpha”) level across the top of the table (columns) and degrees of freedom going in the vertical direction (rows). The value is $\frac{1}{2}$ of the probability level (p). The appropriate column is determined by dividing the p value by 2. For example, if we want to have 80% confidence in the answer, i.e., $p = 0.2$, then α is 0.1.

The number of degrees of freedom is determined by subtracting one from the number of total number of observations (n-1). A soil sample consisting of 21 cores would then have 20 degrees of freedom. So, the t value at the junction of $p = 0.2$, $\alpha = 0.10$, and 20 degrees of freedom is 1.325. A t-table and sample calculations for demonstrating the use of the sampling requirement follow.

Table 33.1. A simplified sample t distribution table. The α value represents the degree of significance desired in the calculations. For a two-sided test the significance level is two times the α value.

Degrees of freedom	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$
1	3.078	6.314	12.706
2	1.886	2.92	4.303
3	1.638	2.353	3.182
4	1.533	2.132	2.776
5	1.476	2.015	2.571
6	1.44	1.943	2.447
7	1.415	1.895	2.365
8	1.397	1.86	2.306
9	1.383	1.833	2.262
10	1.372	1.812	2.228
20	1.325	1.725	2.086
∞	1.282	1.645	1.96

Example 2. Comparison of soil test results

A crop consultant would like to determine the soil test phosphorus (P) value from a field within 2 ppm of the true value 80% of the time. The manager knows that that in a previous field where he collected 21 samples, the variance (s^2) of these samples was 25. The t value of 1.325 (Table 33.1) was calculated in the previous paragraph. Once the t , s^2 , and D values are determined, they are substituted into the subsampling requirement equation. [http://www.ipni.net/ipniweb/insights.nsf/\\$webcontents/A7D980BC92E6385085257810005806EF/\\$file/Mikkelsen+Math+Insights_2010_pr9.pdf](http://www.ipni.net/ipniweb/insights.nsf/$webcontents/A7D980BC92E6385085257810005806EF/$file/Mikkelsen+Math+Insights_2010_pr9.pdf)

$$\alpha = (1 - 0.8) / 2 = 0.10$$

$$t_{0.10, df=20} = 1.325$$

$$n = \frac{t^2 \cdot s^2}{D^2} = \frac{1.325^2 \cdot 25}{2^2} = 10.97 \approx 11$$

This analysis indicates that 11 individual cores should be composited into a single composite sample.

Example 3. Corn hybrid yield

Use Microsoft Excel to determine if the yields from two corn hybrids are the same or dissimilar. Yield data was randomly collected from different fields in a county.

1. Type data into Excel.
2. Select Tools.
3. Select Data Analysis.
4. Select t-Test Two-Sample Assuming Equal Variances, OK.
5. Highlight Data for Variable 1 and 2, select OK.
6. Results are shown below.

Samples	Variety 1	Variety 2
1	110	132
2	100	133
3	99	154
4	75	99
5	47	88
6	112	120
7	---	243

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	Variable 2
Mean	90.5	138.4286
Variance	627.5	2614.286
Observations	6	7
Pooled Variance	1711.201	---
Hypothesized Mean Difference	0	---
Df	11	---
t Stat	-2.08256	---
P(T<=t) one-tail	0.030713	---
t Critical one-tail	1.795885	---
P(T<=t) two-tail	0.061426	---
t Critical two-tail	2.200985	---

For this analysis, a two-tailed test was used. Based on a P(T<=t) two-tail value of 0.061 (1-0.061 = .939= 93.9%) indicates that there is a 93.9% probability that the treatments are different.

Example 4. Determining the highest yielding cultivar

In scientific studies, one, two, or more factors are typically chosen for comparison. In this example, six corn hybrids are compared. Each hybrid is randomly planted in strips in four fields. To solve this problem, an approach for determining the yields of each strip is needed. Yields from strips can be measured with a weigh wagon or with a yield monitor.

If a yield monitor is used, we suggest that you contact an Extension Specialist for assistance. Programs for processing yield monitor data are available at Pierce and Clay (2007). Using a weigh wagon, yield was taken from the harvested strips. The yields from the four replications (the four different fields) were corrected to 15.5% moisture and input into a spreadsheet as shown in Figure 33.5.

	A	B	C	D	E	F	G
1							
2		Variety 1	Variety 2	Variety 3	Variety 4	Variety 5	Variety 6
3	Field 1	126	144	131	150	160	155
4	Field 2	132	149	136	155	158	150
5	Field 3	137	151	144	153	156	158
6	Field 4	155	149	143	162	162	153

Figure 33.5. Corn yield data from a hypothetical on-farm study that evaluated six different hybrids.

In this analysis, each field will be considered as a block. A block contains all of the treatments. The advantage of using a blocked design is that differences between fields can be removed from the analysis. In this example, the first factor is the treatment and the second factor is blocks. To conduct the analysis, select Tools and ANOVA: Two-Factor Without Replication (Fig. 33.6).

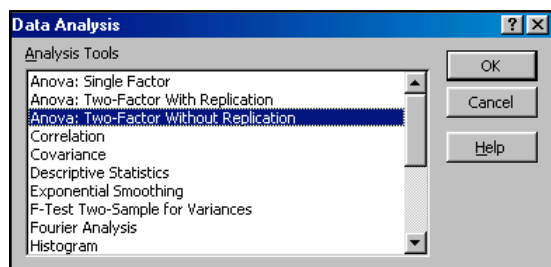


Figure 33.6. Selecting the statistical program that will be used in Example 3.

In the ANOVA Two-Factor Without Replication popup, put the cell numbers into the input range (A2:G6 in this example). Labels are checked (Fig. 33.7). The alpha: level that is default is 0.05 or the 95% probability level is accepted by doing nothing. An Output Range A10 (\$A\$10) is selected (Fig. 33.7). Ok is then entered.

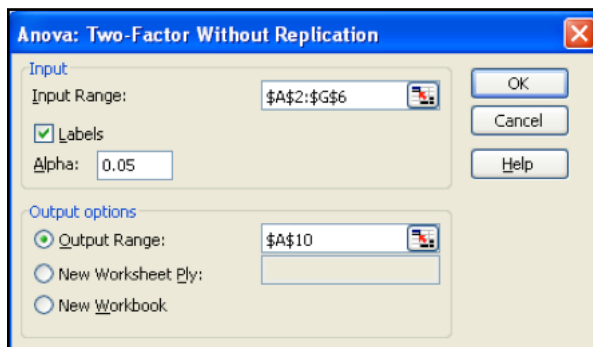


Figure 33.7. Conducting the statistical analysis discussed in Example 3.

Results of the ANOVA are shown in Figure 33.8. The table indicates that the Columns (the different varieties) have an F value of 11.3866 (cell E29). This is higher than the critical F value of 2.901295 (cell G29). This means that at the 0.05 (or the 95% probability level), there are yield differences between varieties.

In fact, the Probability level (P-value) of 0.000112 (cell F29) indicated that the varieties are different at above the 99% ($1 - 0.000112 = 0.999888$ which = 99.9888%) level. This is clearly a highly significant difference. Note that the rows (fields) are also different (cells E28:G28). These cells indicate that the fields (as we might expect) are statistically significantly different (at the greater than 95% level, actually $1 - 0.03542 = 96.458\%$) also.

	A	B	C	D	E
10	Anova: Two-Factor Without Replication				
11					
12	SUMMARY	Count	Sum	Average	Variance
13	Field 1	6	866	144.3333	181.0667
14	Field 2	6	880	146.6667	108.6667
15	Field 3	6	899	149.8333	62.96667
16	Field 4	6	924	154	55.2
17					
18	Variety 1	4	550	137.5	156.3333
19	Variety 2	4	593	148.25	8.916667
20	Variety 3	4	554	138.5	37.66667
21	Variety 4	4	620	155	26
22	Variety 5	4	636	159	6.666667
23	Variety 6	4	616	154	11.33333

Figure 33.8. Statistical results from the hypothetical on-farm corn hybrid study. The p-values indicated that there is a difference between the hybrids.

	A	B	C	D	E	F	G
26	ANOVA						
27	Source of Variation	SS	df	MS	F	P-value	F crit
28	Rows	315.4583	3	105.1528	3.708729	0.03542	3.287383
29	Columns	1614.208	5	322.8417	11.3866	0.000112	2.901295
30	Error	425.2917	15	28.35278			
31							
32	Total	2354.958	23				

An LSD (least significant difference) can be used to determine if the yields from one hybrid is higher or lower than another. To calculate the LSD value, the critical t value must be determined. This is determined using the command, **=TINV(0.05, degrees of freedom)**. The 0.05 value is the significance level, meaning that we want to be 95% certain that there is a difference between the treatments.

An alternative approach is to look this value up in a t-table (Fig 33.1). In this example, the degrees of freedom is provided in cell C30 (Fig. 33.8) and the mean square error, which is similar to the variance discussed above, is provided in cell D30. The number of replications, mean square error, and t-value are needed to calculate the LSD (Fig. 33.9).

	A	B
34	critical t	=TINV(0.05,C30)
35	LSD (.05)	=B34*SQRT(2*D30/4)

	A	B
34	critical t	2.13145
35	LSD (.05)	8.025237

Figure 33.9. Critical t value and LSD calculation. This analysis indicates that if the difference between two hybrids is greater than 8.02 bu/acre, then the yields from the two hybrids are different.

Hybrid 1 averaged 137.5 bu/acre and hybrid 4 averaged 155 bu/ac. The difference between these two hybrids was 17.5 bushels. Hybrid 4 out yielded hybrid 1 because the difference is greater than the LSD value of 8.02 bu/acre (Cell B35 in Fig.33.9).

Summary

This chapter provides an introduction into on-farm research with examples of how to set up experiments and analyze the resulting data. On-farm research can and should be used to test and evaluate management changes on your farm.

Additional information and references

- Clay, D.E., S.A. Clay, C.G. Carlson, and S. Murrell. 2010. Mathematics and Calculations for Agronomists and Soil Scientists. International Plant Nutrition Institute. Available at <http://ppi-store.stores.yahoo.net/maandcaforag.html>
- Clay D.E. and J.F. Shanahan (eds). 2010. GIS Applications in Agriculture: Nutrient Management for Improved Energy Efficiency. Volume 2. Available at <http://www.crcpress.com/product/isbn/0849375266>
- Davis, B. Introduction to Agricultural Statistics. 1999. Delmar Cengage Learning.
- Doerge, T.A., and D.L. Gardner. 1999. The Pioneer Split-planter Comparison Method. In Clay et al. (eds). Site specific management guidelines. Available at [http://www.ipni.net/ppiweb/ppibase.nsf/\\$webindex/article=9C5B38DE85256965005E6A664E4F30BC](http://www.ipni.net/ppiweb/ppibase.nsf/$webindex/article=9C5B38DE85256965005E6A664E4F30BC)
- Pierce F.J., and D.E. Clay. 2007. GIS Applications in Agriculture. 2007. CRC Press, Taylor and Francis, New York, NY. Available at <http://www.crcpress.com/product/isbn/0849375266>
- Knighton, R.E. 1999. Setting up On-farm Experiments. In Clay et al. (eds). Site specific management guidelines. Available at [http://www.ipni.net/ppiweb/ppibase.nsf/\\$webindex/article=3DD9EB2C85256966006141B759745B04](http://www.ipni.net/ppiweb/ppibase.nsf/$webindex/article=3DD9EB2C85256966006141B759745B04)
- Steel, R.D.G., J.H. Torrie, and D.A. Dickey. 1997. Principles and Procedures of Statistics: A Biometrical Approach. Third Edition, McGraw-Hill Publisher.
- Wittig T.A. and Z.W. Wicks III. 1999. Simple On-farm Comparisons. In Clay et al. (eds). Site specific management guidelines. Available at [http://www.ipni.net/ppiweb/ppibase.nsf/\\$webindex/article=F7F42D9E852569660061A77FDDBC70D3](http://www.ipni.net/ppiweb/ppibase.nsf/$webindex/article=F7F42D9E852569660061A77FDDBC70D3)

Acknowledgements

Support for this document was provided by South Dakota State University, South Dakota Corn Utilization Council, USDA-NIFA, South Dakota 2010 research program, South Dakota Wheat Commission, and South Dakota Soybean Board.

Carlson, C.G, C. Ren, and D.E. Clay. 2012. On-farm research and analysis. In Clay, D.E., C.G. Carlson, and K. Dalsted (eds). iGrow Wheat: Best Management Practices for Wheat Production. South Dakota State University, SDSU Extension, Brookings, SD.

In accordance with Federal civil rights law and U.S. Department of Agriculture (USDA) civil rights regulations and policies, the USDA, its Agencies, offices, and employees, and institutions participating in or administering USDA programs are prohibited from discriminating based on race, color, national origin, religion, sex, gender identity (including gender expression), sexual orientation, disability, age, marital status, family/parental status, income derived from a public assistance program, political beliefs, or reprisal or retaliation for prior civil rights activity, in any program or activity conducted or funded by USDA (not all bases apply to all programs). Remedies and complaint filing deadlines vary by program or incident.

Persons with disabilities who require alternative means of communication for program information (e.g., Braille, large print, audiotope, American Sign Language, etc.) should contact the responsible Agency or USDA's TARGET Center at (202) 720-2600 (voice and TTY) or contact USDA through the Federal Relay Service at (800) 877-8339. Additionally, program information may be made available in languages other than English.

To file a program discrimination complaint, complete the USDA Program Discrimination Complaint Form, AD-3027, found online at http://www.ascr.usda.gov/complaint_filing_cust.html and at any USDA office or write a letter addressed to USDA and provide in the letter all of the information requested in the form. To request a copy of the complaint form, call (866) 632-9992. Submit your completed form or letter to USDA by:

(1) mail: U.S. Department of Agriculture
Office of the Assistant Secretary for Civil Rights
1400 Independence Avenue, SW
Washington, D.C. 20250-9410;

(2) fax: (202) 690-7442; or

(3) email: program.intake@usda.gov.

USDA is an equal opportunity provider, employer, and lender.

SDSU Extension is an equal opportunity provider and employer in accordance with the nondiscrimination policies of South Dakota State University, the South Dakota Board of Regents and the United States Department of Agriculture.